Domain Samples Shifting with Auto Encoder Network for Unsupervised Domain Adaptation

Meina Kan

Co-authors: Shiguang Shan, Junting Wu, Xilin Chen

Institute of Computing Technology of the Chinese Academy of Sciences

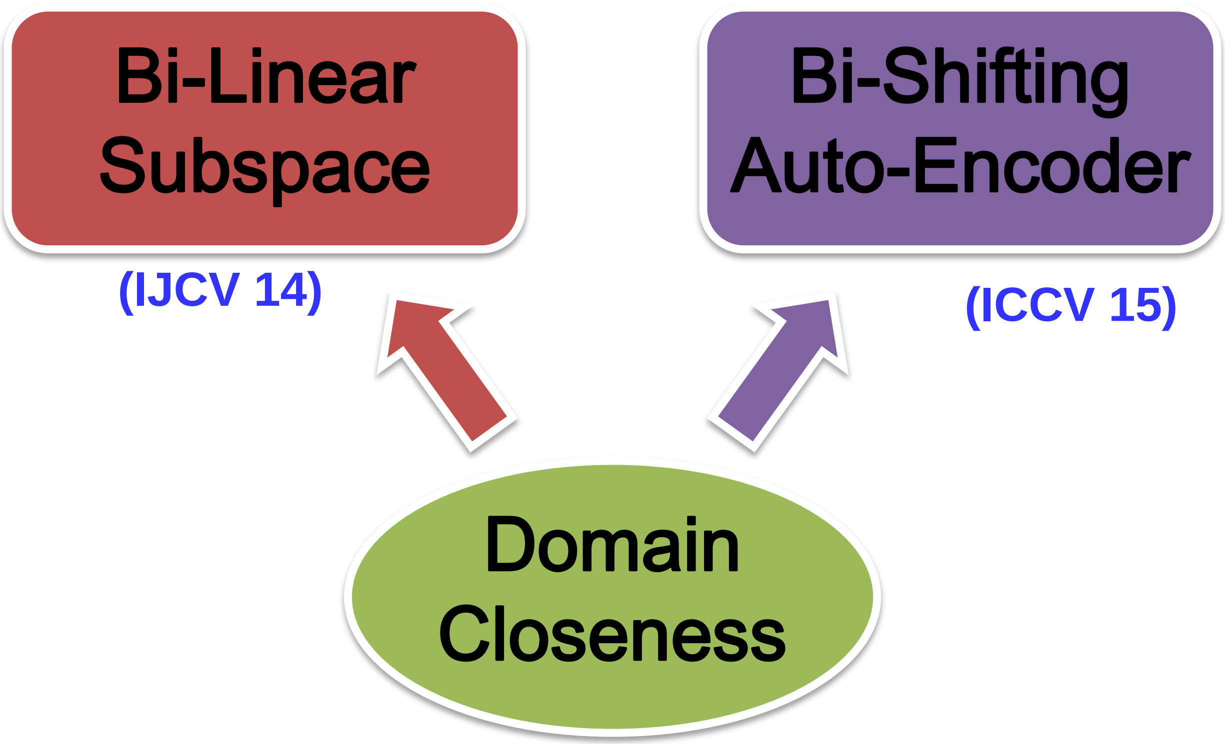
kanmeina@ict.ac.cn, http://vipl.ict.ac.cn/members/mnkan



Institute of Computing Technology,Chinese Academy of Sciences

Outline

- Domain Adaptation for Face Recognition



- Meina Kan, Junting Wu, Shiguang Shan, Xilin Chen. Domain Adaptation for Face Recognition: Targetize Source Domain Bridged by Common Subspace. IJCV, 2014.
- Meina Kan, Shiguang Shan, Xilin Chen. Bi-shifting Auto-Encoder for Unsupervised Domain Adaptation. International Conference on Computer Vision (ICCV), 2015.

Related Works

- Unsupervised Domain Adaptation
  - Data: Labeled Source Domain + Unlabeled Target Domain
  - Task: Recognition on Target Domain
  - Challenge: distribution discrepancy → performance degeneration
- Solutions
  - Reduce the domain discrepancy
    - Instance, feature, or model

Related Works

- Instance-based approaches
  - Resample/reweight source domain
    - Sample Selection Bias/ Covariant Shift[1]
    - Sample selection bias: $p_1(x) \neq p_2(x), p_1(y|x) \neq p_2(y|x)$
    - Covariant Shift: $p_1(x) \neq p_2(x), p_1(y|x) = p_2(y|x)$
  - Criterion: equalize the distribution of each sample

1. Quionero-Candela, et al, Data Shift in Machine Learning, MIT Press 2009

Related Works

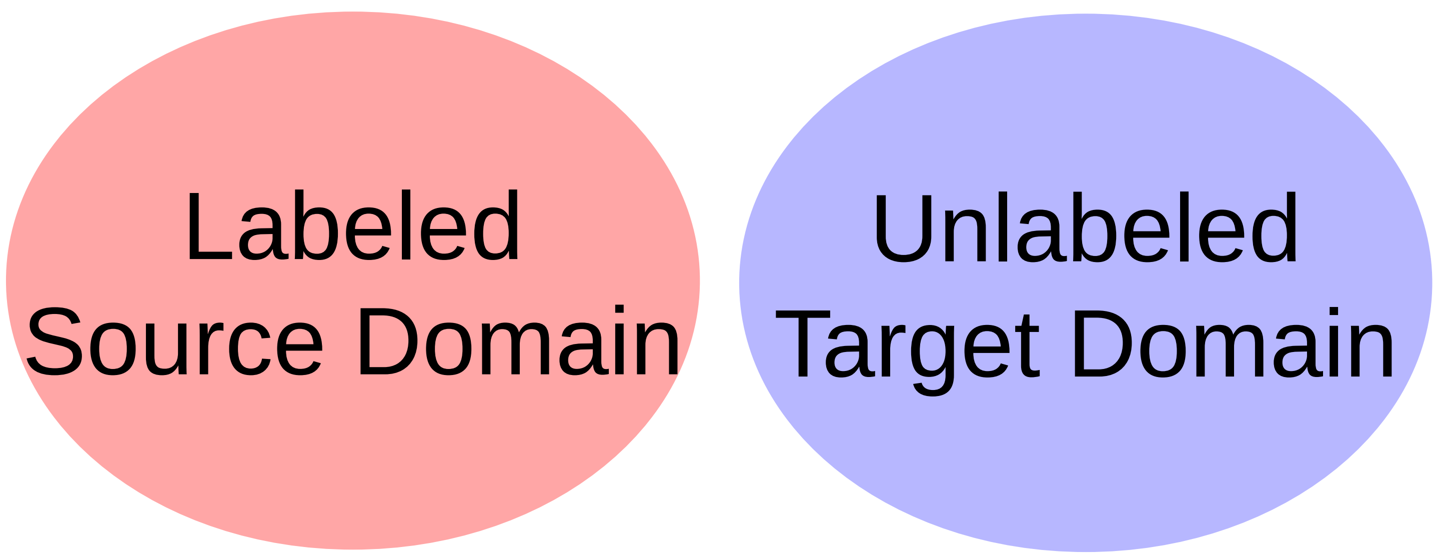
- Feature-based approaches
  - Feature that minimize the discrepancy
    - structural correspondence learning [Blitzer etal., 2006]
    - Transfer Component Analysis [Pan etal., 2009]
    - Sampling Geodesic Flow [Gopalan etal., 2011]
    - Geodesic Flow Kernel [Gong etal., CVPR 2011]
  - Criteria
    - maximum mean discrepancy
    - K-L divergence
    - Bregman divergence
    - Low-rank

Related Works

- Model-based approaches
  - Transductive SVM [Chen et al. PRL 2003]
  - Domain Adaption SVM [Bruzzone etal. PAMI 2010]
  - Addaptive AML[Duan, etal. TPMAI 2012]
  - Domain Transfer SVM [L. Duan 09]
- Criteria
  - Distribution distance
  - Iteratively include more target domain samples

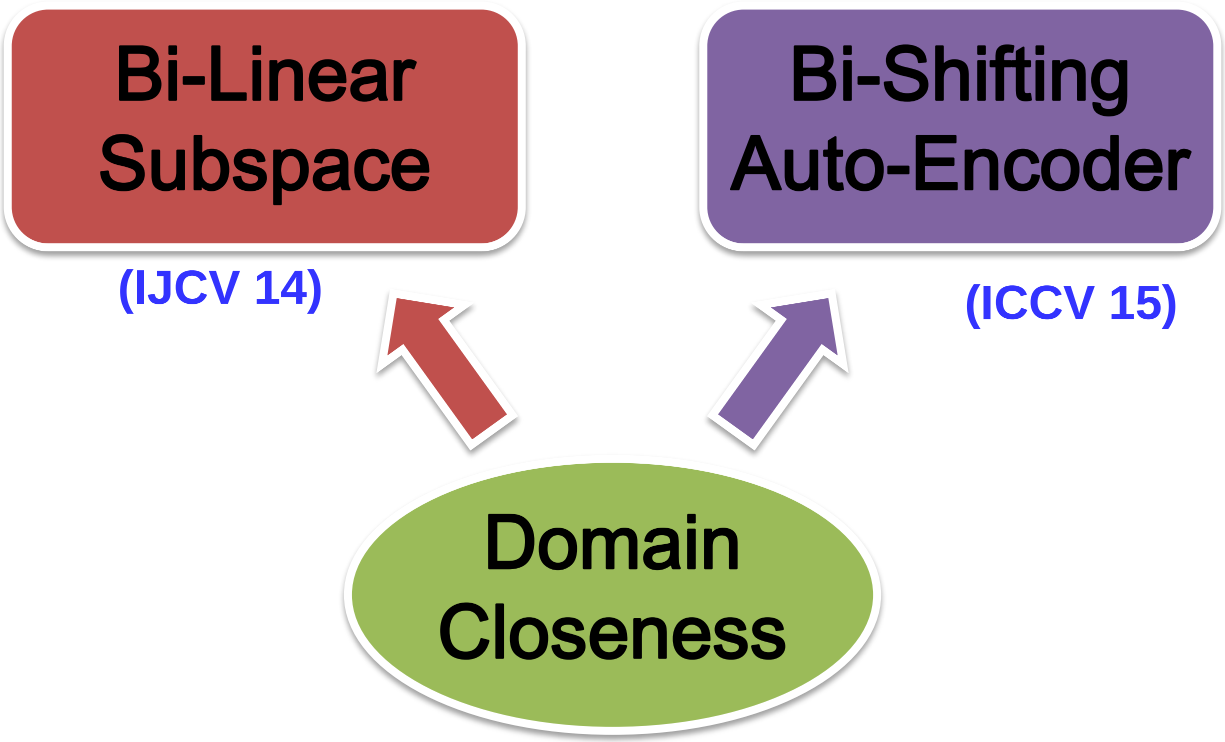
Our solutions

- Instance-based approaches
  - Targetize Source Domain
    - Transform source domains samples to target domain



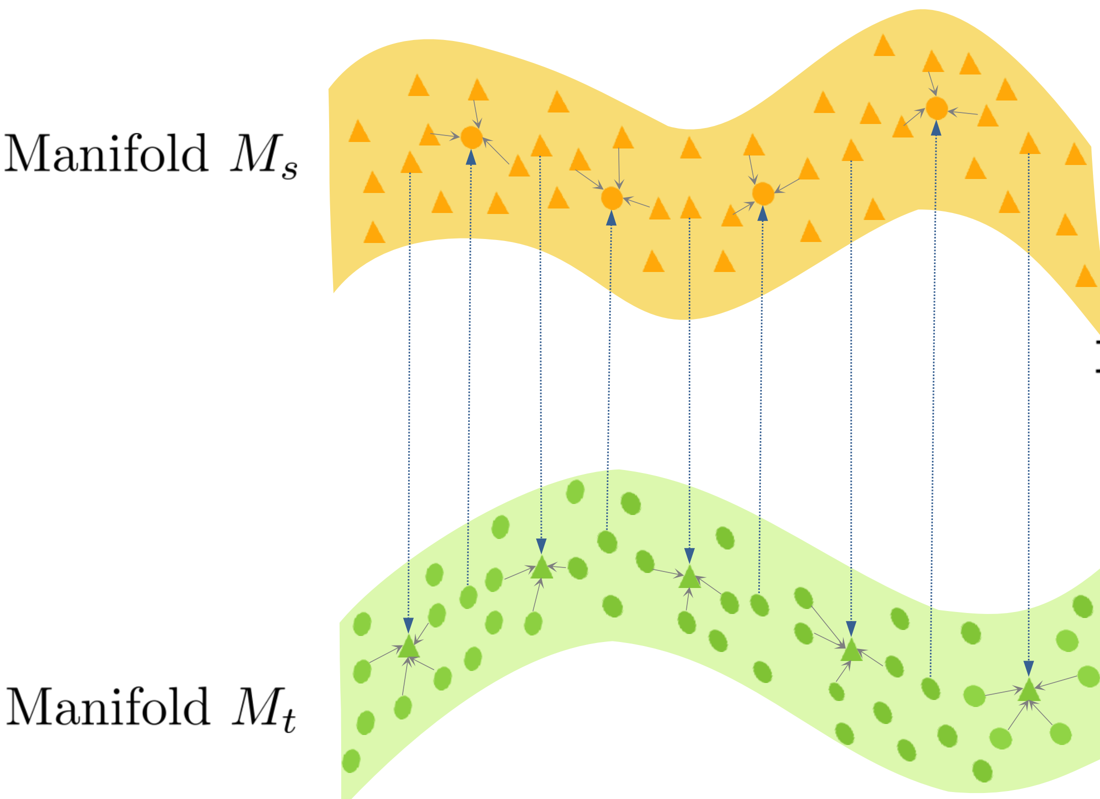
Outline

- Domain Adaptation for Face Recognition



Part I: Sparse Reconstruction for domain closeness

- Why sparse reconstruction for domain closeness?
  - Distance between corresponding local structure



$$\sum_{i=1}^{n_s} ||\mathbf{x}_i^s - \mathbf{x}_i^{t*}||^2$$

$$\mathbf{x}_i^{t*} = \sum_{k=1}^{r} b_k \mathbf{x}_{ik}^t = \mathbf{X}_t \beta_i^t$$

$$\downarrow$$

$$\min \sum_{i=1}^{n_s} ||\mathbf{x}_i^s - \mathbf{X}_t \beta_i^t||^2$$

$\mathbf{x}_i^{t*}|_{i=1}^{n_s}$: a sampling of target domain samples

Part I: Sparse Reconstruction for domain closeness

- Why sparse reconstruction for domain closeness?
  - **Distance between the distributions** of two domains can be approximated by **distance between local structures**

$$\sum_{i=1}^{n_s} \underbrace{||\mathbf{x}_i^s - \mathbf{X}_t \beta_i^t||^2}_{\text{Distance of this sample to target domain}} + \sum_{i=1}^{n_t} \underbrace{||\mathbf{x}_i^t - \mathbf{X}_s \beta_i^s||^2}_{\text{Distance of this sample to source domain}}$$

$$s.t. |\beta_i^t|_1 < \tau, |\beta_i^s|_1 < \tau$$

$\mathbf{x}_i^s$: the i-th source domain sample

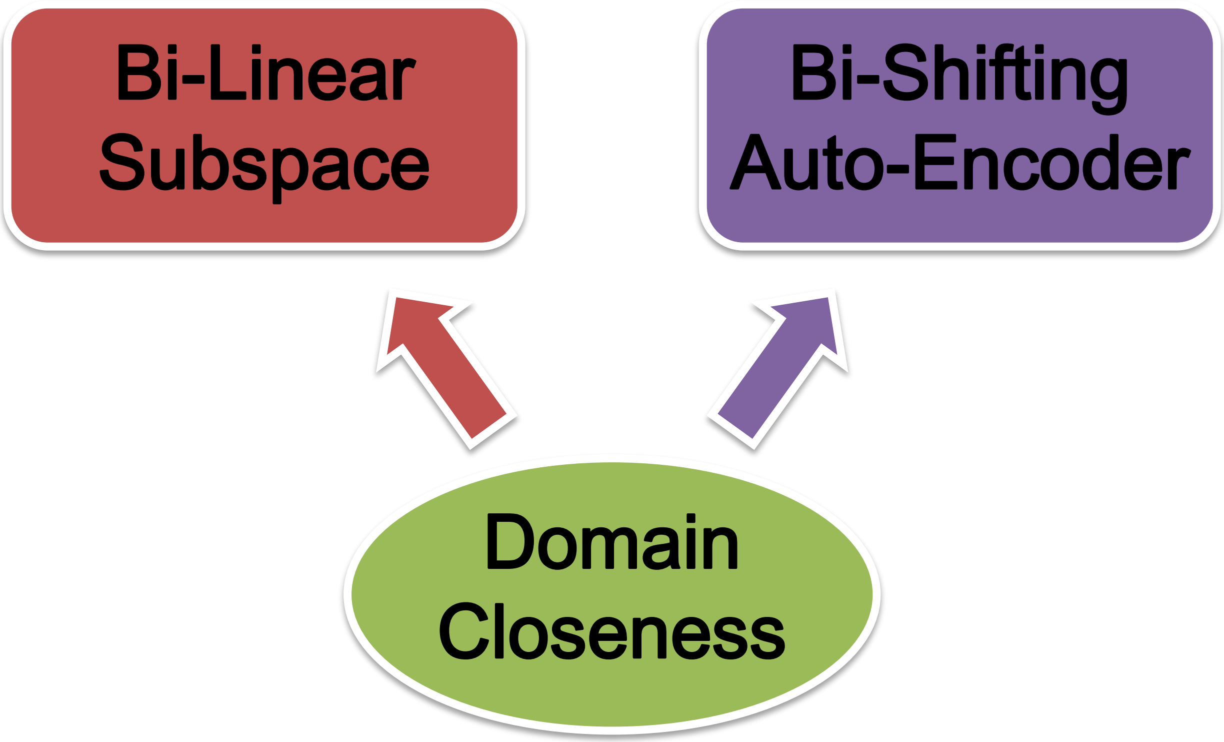
$\mathbf{X}_t$: all target domain samples

Part I: Sparse Reconstruction for domain closeness

- Advantages
  - Non-parametric, flexible to any distribution
    - No need to estimate the distribution
  - Preserve the diverse of shifted domain samples
  - Simple formulation and optimization

Outline

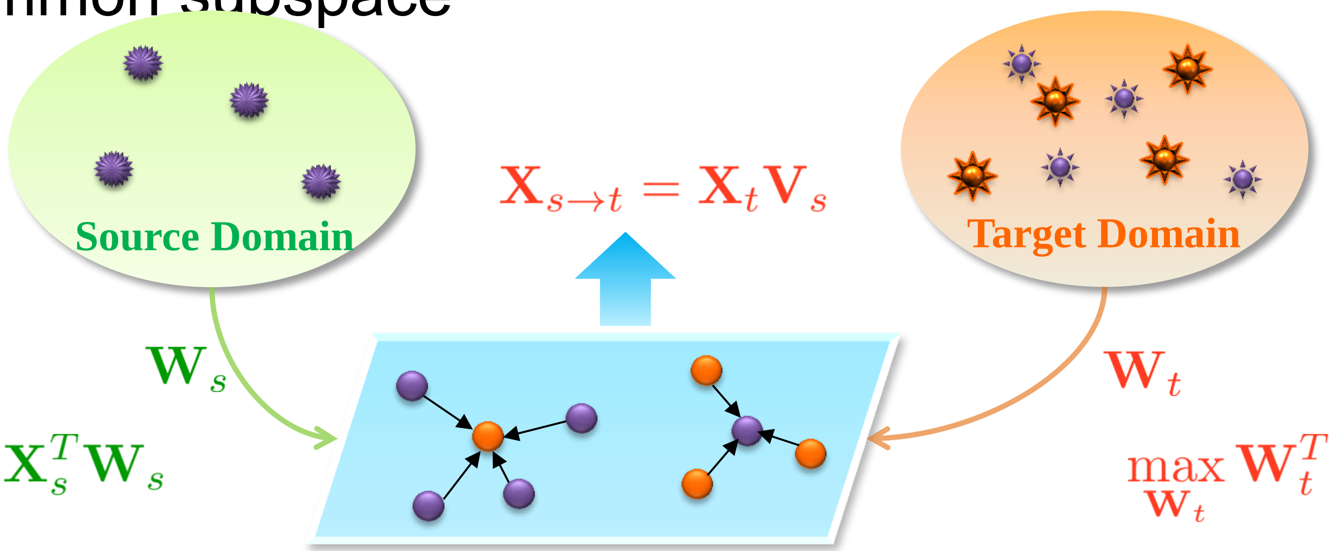
- Domain Adaptation for Face Recognition



- Meina Kan, Junting Wu, Shiguang Shan, Xilin Chen. Domain Adaptation for Face Recognition: Targetize Source Domain Bridged by Common Subspace. IJCV, 2014.

Part II: Bi-linear Subspace for DA

- Basic Idea
  - Transform source domain samples to target domain by employing linear reconstruction in image space
  - The linear reconstruction coefficients are learnt in the common subspace



$$\max_{\mathbf{W}_s} \mathbf{W}_s^T\mathbf{X}_s\mathbf{X}_s^T\mathbf{W}_s \qquad \max_{\mathbf{W}_t} \mathbf{W}_t^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{W}_t$$

$$\min_{\mathbf{V}_s,\mathbf{V}_t} ||\mathbf{W}_s^T\mathbf{X}_s - \mathbf{W}_t^T\mathbf{X}_t\mathbf{V}_s||_2^2 + ||\mathbf{W}_t^T\mathbf{X}_t - \mathbf{W}_s^T\mathbf{X}_s\mathbf{V}_t||_2^2$$

$$\|\mathbf{v}_i^s|_0 < \tau, |v_i^t|_0 < \tau$$

Part II: Bi-linear Subspace for DA

- Formulation
  - Common subspace learning

$$\mathbf{Z}_s = \mathbf{W}_s^T \mathbf{X}_s; \mathbf{Z}_t = \mathbf{W}_t^T \mathbf{X}_t$$

Preserve Structure

$$\max \frac{Tr\left(\frac{1}{n_s}\mathbf{Z}_s\mathbf{Z}_s^T + \frac{1}{n_t}\mathbf{Z}_t\mathbf{Z}_t^T\right)}{\frac{1}{n_s}\|\mathbf{Z}_s - \mathbf{Z}_t\mathbf{V}_s\|_F^2 + \frac{1}{n_t}\|\mathbf{Z}_t - \mathbf{Z}_s\mathbf{V}_t\|_F^2}$$

$$s.t. \ |\mathbf{v}_i^s|_0 \le \tau, |\mathbf{v}_j^t|_0 \le \tau$$

Domain Closeness

- Optimization
  - Alternatively optimize $\mathbf{W}_s, \mathbf{W}_t$ and $\mathbf{V}_s, \mathbf{V}_t$

Part II: Bi-linear Subspace for DA

- Recognition
  - Transform source samples to target domain

In Common Space

$$\mathbf{z}_i^s \approx \mathbf{Z}_t \mathbf{v}_i^s$$

In Image Space

$$\mathbf{x}_i^{s\to t} = \mathbf{X}_t \mathbf{v}_i^s$$

$$\mathbf{z}_i^s = \mathbf{W}_s^T \mathbf{x}_i^s$$

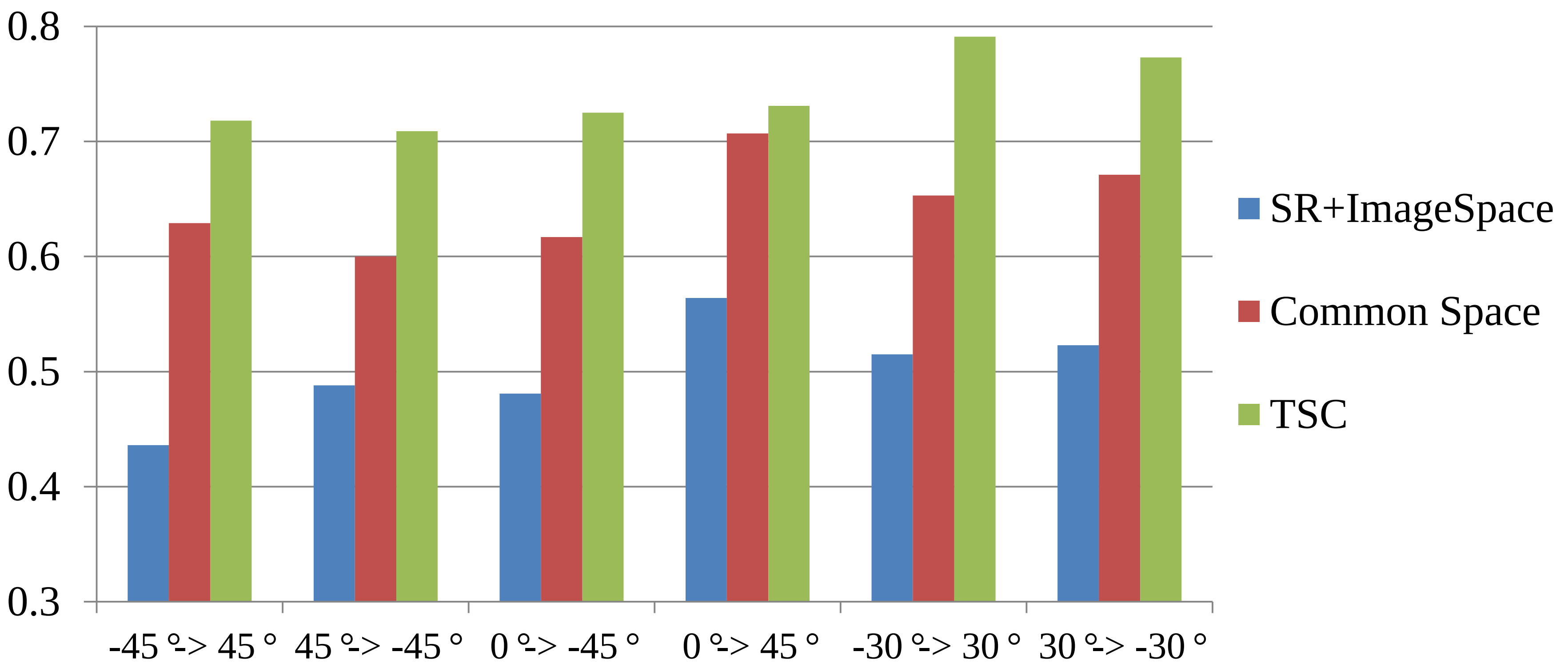
$$\mathbf{z}_i^t = \mathbf{W}_t^T \mathbf{x}_i^t$$

$$\mathbf{x}_i^t \mathbf{v}_i^s = \underbrace{\mathbf{X}_t^{(c)} \mathbf{v}_i^s}_{common} + \underbrace{\left(\mathbf{X}_t - \mathbf{X}_t^{(c)}\right) \mathbf{v}_i^s}_{paticular}$$

  - Classifier for target domain
    - $(\mathbf{X}_{s\to t}, \mathbf{y}_s)$: labeled, similar as target domain
    - Fisher Discriminant Analysis + 1NN classifier

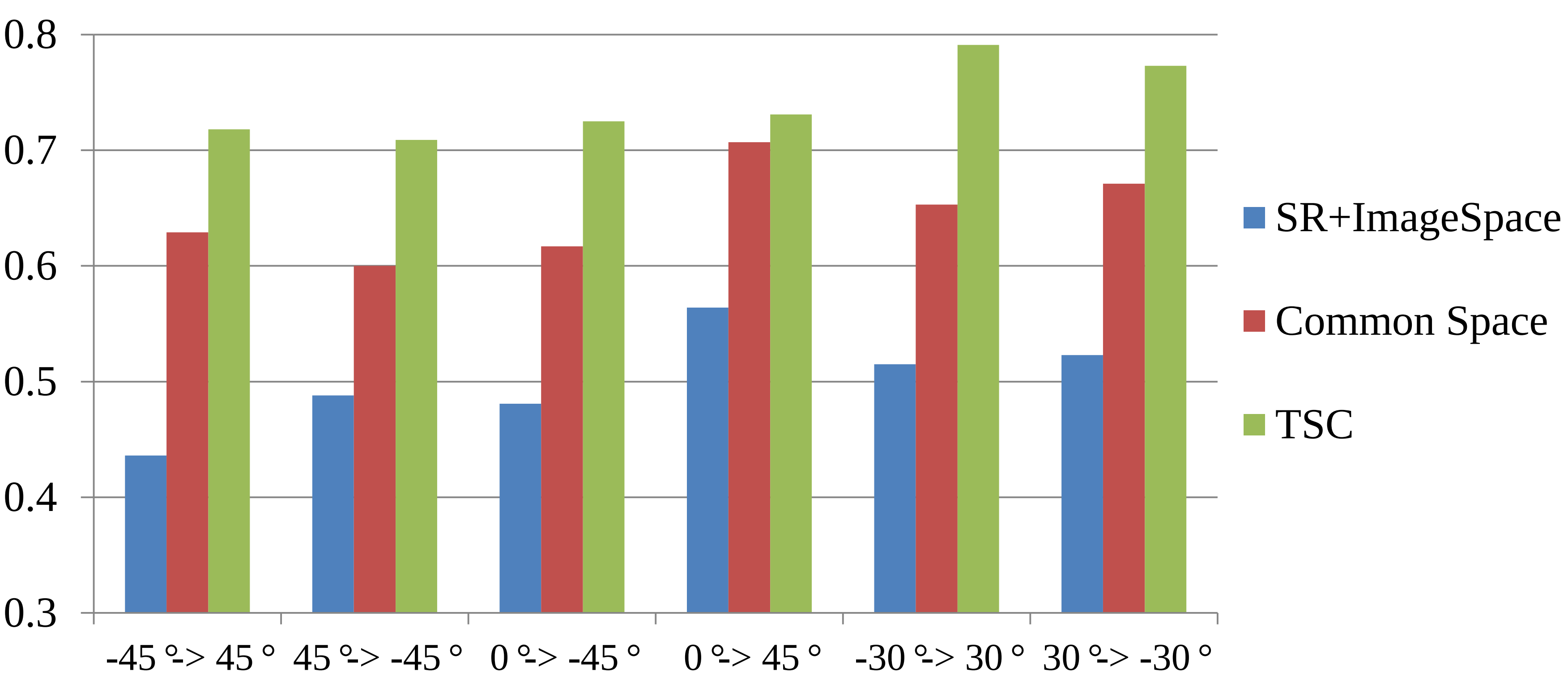
Part II: Bi-linear Subspace for DA

- Experiments: necessity of common space
  - Rank-1 recognition rate
  - Adaptation between poses on MultiPIE



Part II: Bi-linear Subspace for DA

1. **Common Space bridges the domain discrepancy**
2. **Transformation in image space is beneficial**



Part II: Bi-linear Subspace for DA

- Exp: comparison with existing works
  - Rank-1 recognition Rate
  - Adaptation between imaging condition
    - XM2VTS & FRGC

Source	Target	PCA	Na ïve-FLD	TDR	SGF	ITL	TSC
Uncontrol	Control	0.719	0.767	0.831	0.791	0.731	0.855
Control	Uncontrol	0.012	0.083	0.183	0.073	0.041	0.232

Part II: Bi-linear Subspace for DA

- Exp: comparison with existing works
  - Rank-1 recognition Rate
  - Adaptation between ethnicities
    - OFD & XM2VTS

Source	Target	PCA	Naïve-FLD	TDR	SGF	ITL	TSC
Mongolian	Caucasian	0.745	0.768	0.845	0.783	0.715	0.858
Caucasian	Mongolian	0.268	0.593	0.971	0.653	0.11	0.974

Part II: Bi-linear Subspace for DA

- Advantages
  - Common subspace
    - Bridge of two domains, and stability of the reconstruction
  - Shifting in the original image space
    - Preserve information specific to target domain
    - Domain specific information is beneficial

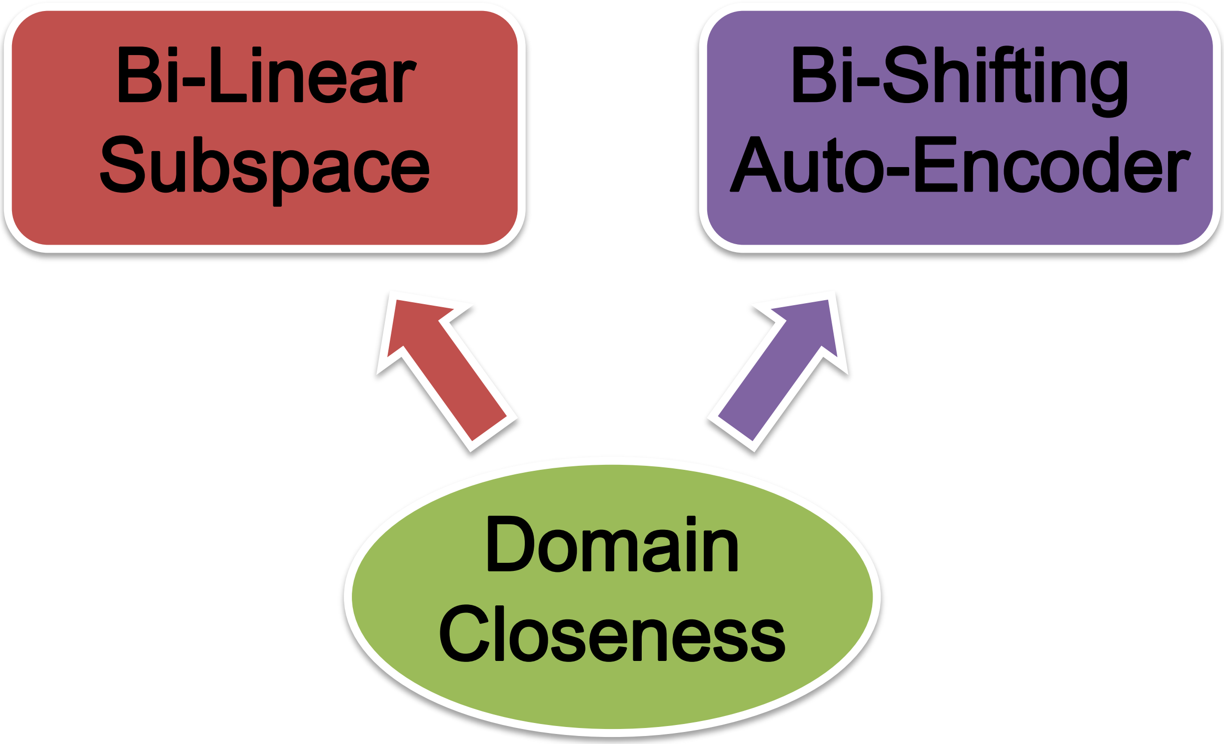
$$\mathbf{X}_{s\rightarrow t} = \mathbf{X}_t\mathbf{V}_s^* = \underbrace{\mathbf{X}_t^{(c)}\mathbf{V}_s^*}_{common} + \underbrace{\left(\mathbf{X}_t - \mathbf{X}_t^{(c)}\right)\mathbf{V}_s^*}_{paticular}$$

The shifting relationship in common space may be inconsistent with that in the image space

Outline

- Domain Adaptation for Face Recognition

$$\mathbf{X}_{s \to t} = \mathbf{X}_t \mathbf{V}_s^*$$



- Meina Kan, Shiguang Shan, Xilin Chen. Bi-shifting Auto-Encoder for Unsupervised Domain Adaptation. International Conference on Computer Vision (ICCV), 2015.

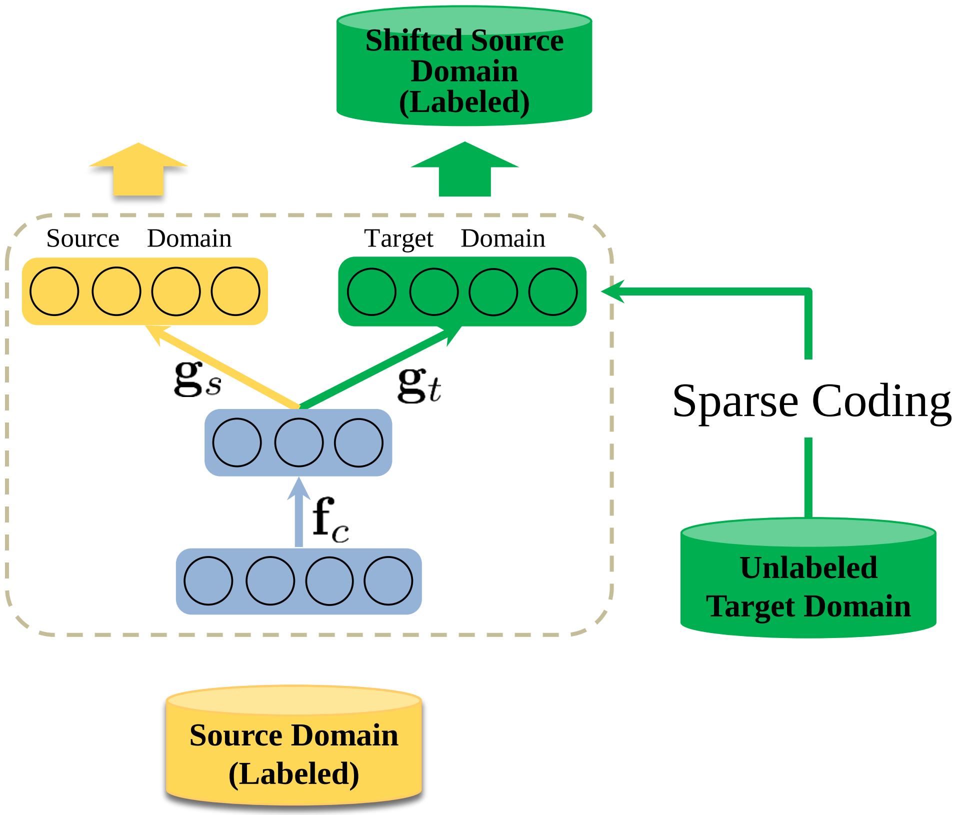
Bi-shifting Auto-Encoder

- Main Idea: shifting samples between domain
- Approach
  - Non-linear model for shifting
    - Shifting directly in the original image space
  - Sparse-reconstruction for closeness measurement
    - Shifted domain and desirable domain

Bi-shifting Auto-Encoder

- Domain Distance
  - ✓ Local spare coding

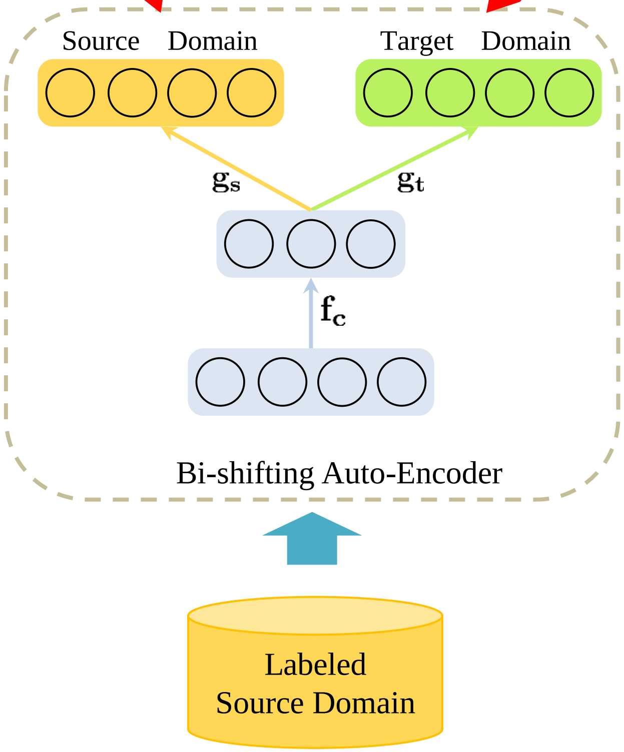
- Shifting model
  - ✓ Non-linear Auto-Encoder



Bi-shifting Auto-Encoder

- Formulation $\min \Omega_1 + \Omega_2$

$$\min \Omega_1 = \min_{\mathbf{f}_c, \mathbf{g}_s, \mathbf{g}_t, \mathbf{B}_s} \underbrace{||\mathbf{X}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_s))||_2^2}_{\text{Reconstruct Itself}} + \underbrace{||\mathbf{X}_t\mathbf{B}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))||_2^2 + \gamma \sum_{i=1}^{n_s} |\beta_i^t|_1}_{\text{Closeness to Target Domain}}$$



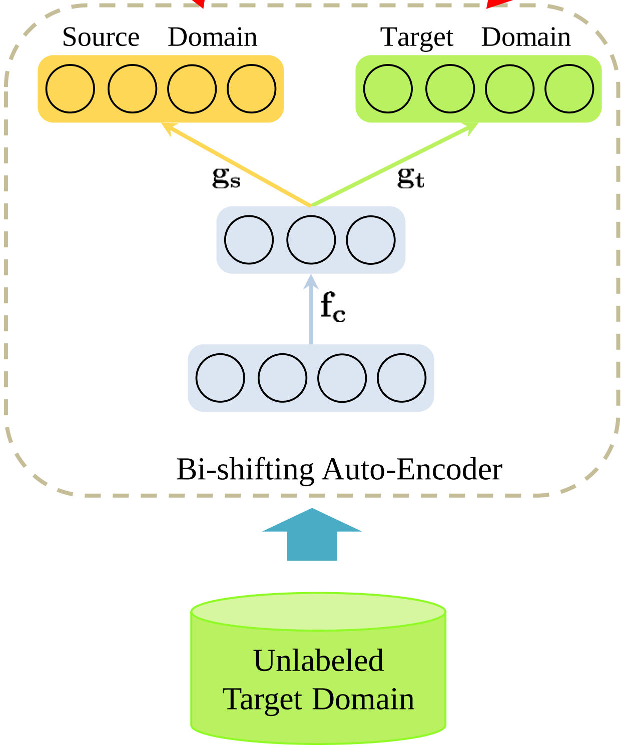
Bi-shifting Auto-Encoder

- Formulation $\quad \min \Omega_1 + \Omega_2$

$$\min \Omega_2 = \min_{\mathbf{f}_c, \mathbf{g}_s, \mathbf{g}_t, \mathbf{B}_s} ||\mathbf{X}_s \mathbf{V}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_t))||_2^2 + ||\mathbf{X}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_t))||_2^2 + \gamma \sum_{i=1}^{n_t} |\boldsymbol{\beta}_i^s|_1$$

Closeness to Target Domain

Reconstruct Itself



Bi-shifting Auto-Encoder

- Formulation

$$\min_{\mathbf{f}_c,\mathbf{g}_s,\mathbf{g}_t,\mathbf{B}_s,\mathbf{B}_t} ||\mathbf{X}_s-\mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_s))||_2^2+||\mathbf{X}_t\mathbf{B}_t-\mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))||_2^2$$
$$+||\mathbf{X}_s\mathbf{B}_s-\mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_t))||_2^2+||\mathbf{X}_t-\mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_t))||_2^2$$
$$+\gamma\left(\sum_{i=1}^{n_s}|\boldsymbol{\beta}_i^t|_1+\sum_{i=1}^{n_t}|\boldsymbol{\beta}_i^s|_1\right).$$

Sparse Constraint

Target Domain Samples

Source Domain Samples

Bi-shifting Auto-Encoder

- Optimization
  - Alternative optimize AE and Sparse Reconst.

Fix AE, Optimize Sparse reconstruction

$$\min_{\mathbf{B}_s,\mathbf{B}_t} ||\mathbf{X}_t\mathbf{B}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))||_2^2 + ||\mathbf{X}_s\mathbf{B}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_t))||_2^2$$
$$+\gamma\left(\sum_{i=1}^{n_s}|\boldsymbol{\beta}_i^t|_1 + \sum_{i=1}^{n_t}|\boldsymbol{\beta}_i^s|_1\right).$$

Least Angel Regression for optimizing Sparse Reconstruction

Fix Sparse reconstruction, Optimize AE

$$\min_{\mathbf{f}_c,\mathbf{g}_s,\mathbf{g}_t} ||\mathbf{X}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_s))||_2^2 + ||\mathbf{X}_t\mathbf{B}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))||_2^2$$
$$+ ||\mathbf{X}_s\mathbf{B}_s - \mathbf{g}_s(\mathbf{f}_c(\mathbf{X}_t))||_2^2 + ||\mathbf{X}_t - \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_t))||_2^2$$

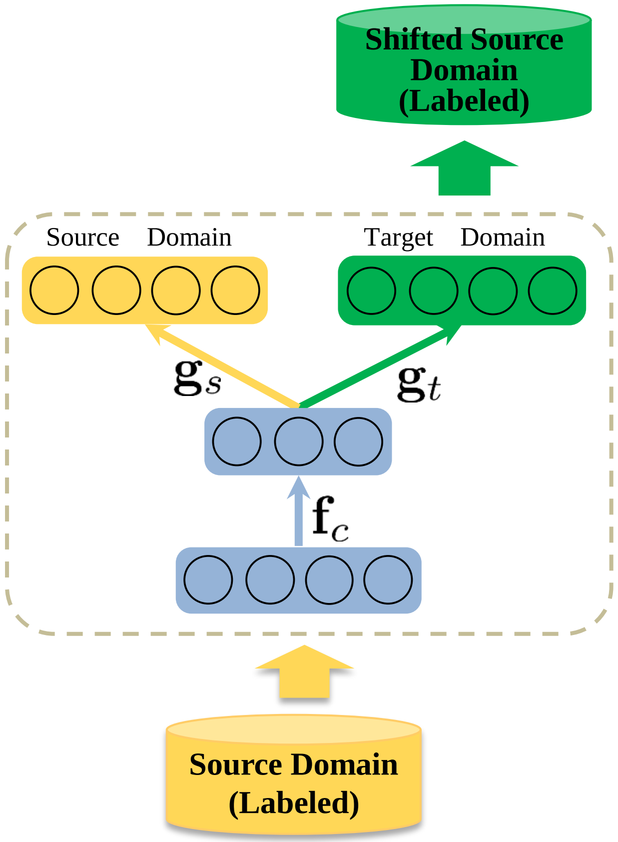
Gradient Descent for optimizing Bi-shifting AE

Bi-shifting Auto-Encoder

- Shift Source Domain
  - Labeled
  - Share similar distribution as target domain

$$\mathbf{G}_t \triangleq \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))$$

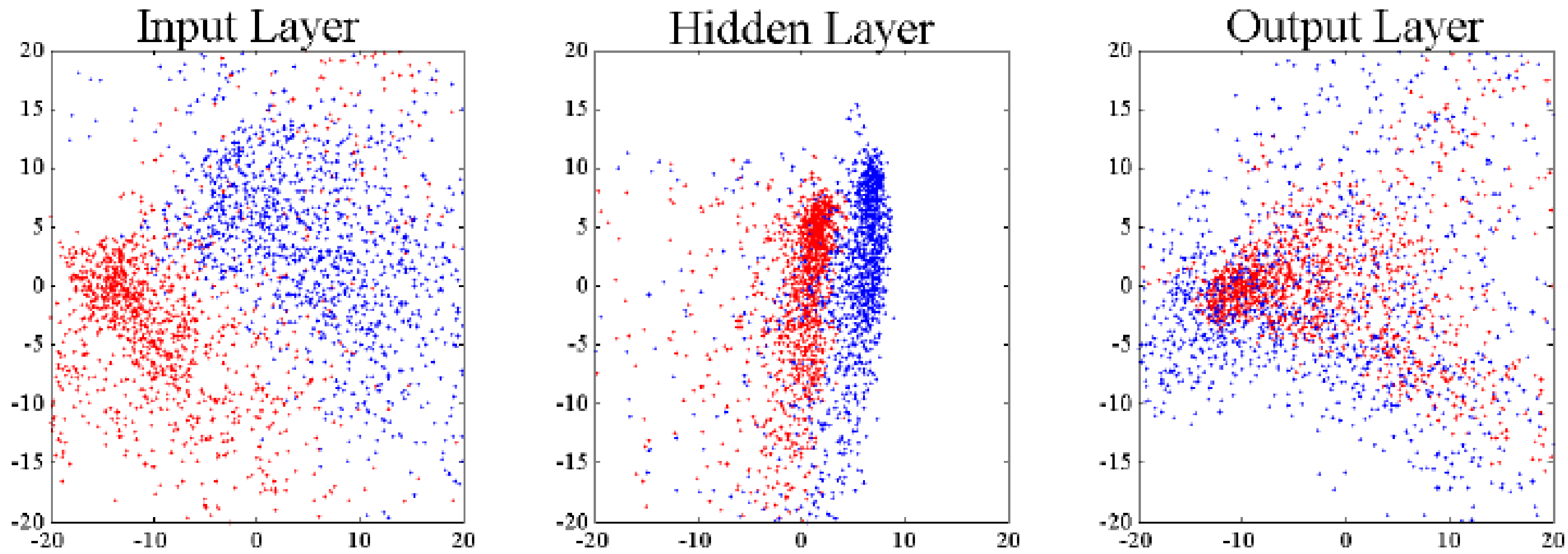
$$(\mathbf{X}_s, \mathbf{y}_s) \to (\mathbf{G}_t, \mathbf{y}_s)$$



- Classifier for target domain
  - Fisher Discriminant Analysis on $(\mathbf{G}_t, \mathbf{y}_s)$
  - 1NN classifier

Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between poses on MultiPIE



Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between poses on MultiPIE

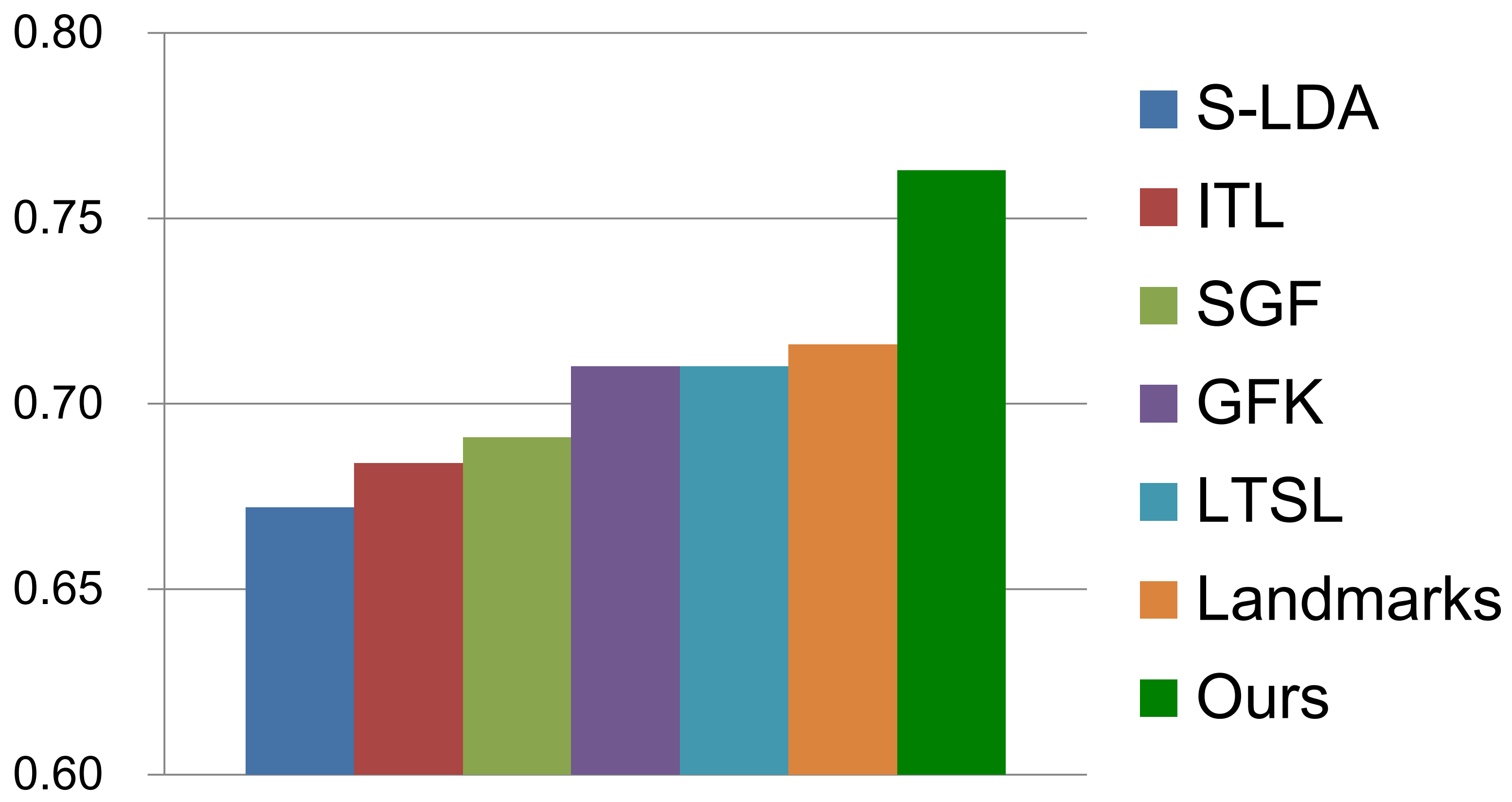
Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between poses on MultiPIE

Methods	-45° →0°	-30° →15°	-15° →30°	0° →45°	-45° →-15°	-15° →15°	15° →45°	Average
Source LDA	0.665	0.693	0.669	0.617	0.703	0.719	0.637	0.672
Target PCA	0.615	0.632	0.583	0.541	0.651	0.632	0.541	0.599
ITL	0.745	0.727	0.653	0.597	0.721	0.714	0.634	0.684
SGF+LDA	0.716	0.714	0.669	0.629	0.735	0.748	0.629	0.691
GFK (PCA,LDA)	0.751	0.754	0.699	0.615	0.767	0.761	0.624	0.710
Landmarks (PCA,LDA)	0.747	0.759	0.701	0.640	0.763	0.763	0.642	0.716
LTSL (LDA)	0.736	0.735	0.698	0.642	0.752	0.767	0.640	0.710
AE+LDA	0.735	0.708	0.702	0.656	0.746	0.739	0.649	0.705
BAE+LDA(Ours)	**0.795**	**0.794**	**0.763**	**0.698**	**0.803**	**0.796**	**0.693**	**0.763**

Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between poses on MultiPIE



Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between ethnicity

Methods	Cau→Mon	Mon→Cau	Average
Source LDA	0.679	0.676	0.678
ITL	0.801	0.775	0.788
SGF+LDA	0.790	0.751	0.771
GFK (PCA,LDA)	0.738	0.721	0.730
Landmarks (PCA,LDA)	0.718	0.763	0.741
LTSL (LDA)	0.791	0.793	0.792
AE+LDA	0.784	0.786	0.785
BAE+LDA(Ours)	**0.892**	**0.826**	**0.859**

Bi-shifting Auto-Encoder

- Experiments
  - Adaptation between lighting

Methods	VIS→NIR	NIR→VIS	Average
Source LDA	0.816	0.779	0.798
ITL	0.858	0.877	0.868
SGF+LDA	0.841	0.832	0.837
GFK (PCA,LDA)	0.850	0.867	0.859
Landmarks (PCA,LDA)	0.859	0.871	0.865
LTSL (LDA)	0.868	0.878	0.873
AE+LDA	0.827	0.846	0.837
BAE+LDA(Ours)	**0.904**	**0.920**	**0.912**

Summary

- Domain Adaptation for Face Recognition

$$\mathbf{X}_{s\rightarrow t} = \mathbf{X}_t \mathbf{V}_s^*$$

Bi-Linear Subspace

$$\mathbf{G}_t \triangleq \mathbf{g}_t(\mathbf{f}_c(\mathbf{X}_s))$$

Bi-Shifting Auto-Encoder

Domain Closeness

$$\min \sum_{i=1}^{n_s} ||\mathbf{x}_i^s - \mathbf{X}_t \beta_i^t||^2$$
$$s.t. |\beta_i^t|_1 < \tau$$

Summary

- Both commonality and particularity are beneficial
- Local Sparse Reconstruction can effectively ensure the domain closeness
  - non-parametric, flexible to complex distribution
- Deep/Non-linear model can better characterize the domain discrepancy
  - makes it feasible to shift samples between domains in image space

Our works about TL on ICCV 2015

- Shifting samples between domains
  - Meina Kan, Shiguang Shan, Xilin Chen. Bi-shifting Auto-Encoder for Unsupervised Domain Adaptation. International Conference on Computer Vision (ICCV), 2015.
- Shifting annotations across datasets
  - Jie Zhang, Meina Kan, Shiguang Shan, Xilin Chen. Deep Regression Network for Face Alignment by Leveraging Datasets with Varying Annotations. International Conference on Computer Vision (ICCV), 2015.
- Generic-to-Specific TL for DCNN pretraining
  - Xin Liu, Shaoxin Li, Meina Kan, Jie Zhang, Shuzhe Wu, Wenxian Liu, hu Han, Shiguang Shan, Xilin Chen. AgeNet: Deeply Learned Regressor and Classifier for Robust Apparent Age Estimation. ChaLearn Looking at People Workshop on ICCV 2015. (2nd Winner of Apparent Age Estimation).

Thanks, Q & A

Meina Kan
Co-author: Shiguang Shan, Junting Wu, Xilin Chen
kanmeina@ict.ac.cn
http://vipl.ict.ac.cn/members/mnkan